HHS Public Access

Author manuscript

*Nat Microbiol*. Author manuscript; available in PMC 2019 November 01.

Published in final edited form as:

*Nat Microbiol*. 2019 May ; 4(5): 876–887. doi:10.1038/s41564-019-0399-4.

# Determinants of Zika Virus Host Tropism Uncovered by Deep Mutational Scanning

**Yin Xiang Setoh**[1,†,*], **Alberto A. Amarilla**[1,†], **Nias Y.G. Peng**[1], **Rebecca E. Griffiths**[2], **Julio Carrera**[3], **Morgan E. Freney**[1], **Eri Nakayama**[5], **Shinya Ogawa**[6], **Daniel Watterson**[1], **Naphak Modhiran**[1], **Faith Elizabeth Nanyonga**[1], **Francisco J. Torres**[1], **Andrii Slonchak**[1], **Parthiban Periasamy**[1], **Natalie A. Prow**[4], **Bing Tang**[4], **Jessica Harrison**[1], **Jody Hobson-Peters**[1], **Thom Cuddihy**[7], **Justin Cooper-White**[2], **Roy A. Hall**[1], **Paul R. Young**[1], **Jason M Mackenzie**[3], **Ernst Wolvetang**[2], **Jesse D. Bloom**[8,9], **Andreas Suhrbier**[4,#], **Alexander A. Khromykh**[1,#,*]

[1]Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Queensland, Australia.

[2]Australian Institute for Bioengineering and Nanotechnology, UQ stemCARE, University of Queensland, St Lucia, Queensland, Australia

[3]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, Australia.

[4]Inflammation Biology Group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

[5]Department of Virology I, National Institute of Infectious Diseases, Tokyo, Japan.

[6]Department of Applied Biological Chemistry, School of Agriculture and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan.

[7]Research Computing Centre, The University of Queensland, St. Lucia, Queensland, Australia.

[8]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, United States.

[9]Howard Hughes Medical Institute, Seattle, United States

## Abstract

Arboviruses cycle between, and replicate in, both invertebrate and vertebrate hosts, which for Zika virus (ZIKV) involves *Aedes* mosquitoes and primates[1]. The viral determinants required for

[*]Correspondence to: Alexander A. Khromykh alexander.khromykh@uq.edu.au or Yin Xiang Setoh y.setoh@uq.edu.au.

Author contributions:

Conceptualization (YXS, AAK), Funding acquisition (AAK, ASuh, YXS), Performed experiments (YXS, AAA, NYGP, REG, JC, EN, SO, DW, NM, MEF, BT, ASlon, FJT, FEN, PP), Analysis of data (YXS, AAA, REG, JH, EW, SO, JH, AAK, JDB, JMM), Software development (JDB), Establishing computing resources (TC), Provision of critical reagents and models (JHP, RAH, NAP, REG, JMM, EW, PRY, JCW), Writing – original draft (YXS, AAK), Writing – review & editing (YXS, AAK, ASuh, DW, JDB, EW, RAH, AAA).

[†]These authors contributed equally and should be considered as joint first authors.

[#]These authors contributed equally and should be considered as joint senior authors

Competing interests:

Authors declare no competing interests.

replication in such obligate hosts are under strong purifying selection during natural virus evolution, making it challenging to resolve which determinants are optimal for viral fitness in each host. Herein we describe a deep mutational scanning (DMS) strategy[2–5] whereby a viral cDNA library was constructed containing all codon substitutions in the C-terminal 204 amino acids of ZIKV envelope (E) protein. The cDNA library was transfected into C6/36 (*Aedes*) and Vero (primate) cells, with subsequent deep sequencing and computational analyses of recovered viruses showing that K316Q and S461G, or Q350L and T397S substitutions conferred substantial replicative advantages in mosquito and primate cells, respectively. A 316Q/461G virus was constructed and shown to be replication-defective in mammalian cells due to severely compromised virus particle formation and secretion. The 316Q/461G virus was also highly attenuated in human brain organoids, and illustrated utility as a vaccine in mice. This approach can thus imitate evolutionary selection in a matter of days and identify amino acids key to regulating virus replication in specific host environments.

Zika virus (ZIKV) belongs to the family *Flaviviridae* (see Supplementary Notes) and is an emerging mosquito-transmitted human pathogen of significant medical importance[6,7], due to its association with a range of primarily neurological congenital abnormalities now termed congenital Zika syndrome[8]. The viral determinants required for replication in different host environments are under strong purifying selection during natural virus evolution, making it challenging to identify the determinants that are necessary for viral fitness in different hosts. Deep mutational scanning (DMS) of viral proteins is emerging as a powerful tool to address this issue allowing identification of amino acid residues that are optimal for viral replication in specific selective environments in a matter of days[2–5,9,10]. DMS involves generation of viral libraries containing all codon substitutions of a viral protein, selecting these viral libraries in different host environments, and employing deep sequencing and computational analyses to identify substitutions that are selected in the specific host environments. Herein we applied DMS to the ZIKV E protein with the aim of identifying amino acids beneficial for virus replication in mammalian or mosquito host cells.

## Results

### Deep mutational scanning (DMS) of C-terminal 204 residues of ZIKV E protein and selection of beneficial mutations in mosquito and mammalian cells.

We performed DMS of the C-terminal 204 residues of ZIKV (Natal 2015 strain[11,12]) E protein, which encompasses the putative receptor-binding domain III (E-DIII) and the stem-anchor regions (E-TM1 and E-TM2). This region of E protein was selected (i) because of its role in mediating receptor-dependent infection and (ii) to avoid the N-terminal part of E, which contains an N-glycosylation site at position 154. Previous studies showed that ablation of glycosylation at this site improved ZIKV infection in mosquito cells[13,14].

Using DMS primers (Supplementary Table 1) encoding a central codon-targeted “NNN” sequence (N=A/T/G/C), a cDNA amplicon library was generated (Fig 1a) in which each of the C-terminal 204 codons was designed to be substituted with all possible codon and amino acid variants (there are 204 × 63 = 12,852 possible single codon substitutions, or 204 × 19 = 3,876 possible single amino acid substitutions). To minimize sequencing errors, a barcoded

subamplicon method[2–5] was used (see Methods). Computational analysis of deep sequencing data using *dms_tools2*[15] (Supplementary Data 1, see also https://github.com/jbloomlab/ZIKV_DMS_w_Khromykh) identified 12,200 codon variants coding 3,827 amino acid variants in the amplicon library, representing 94.9% and 98.7% of all possible codon or amino acid variants, respectively (Fig 1b and c). Both synonymous and non-synonymous substitutions were present in the CPER cDNA amplicon library, with non-synonymous mutations dominating (Supplementary Figure 1a). More than 90% of possible single codon variants and >95% of possible single amino acid variants were represented at least once at every position in the library (Supplementary Figure 1b). The amplicon library was included into a circular polymerase extension reaction (CPER) assembly with other ZIKV cDNA fragments to generate an infectious full-length ZIKV DMS cDNA library[12,16–20]. This cDNA library was then transfected into Vero or C6/36 cells and transfected cells were incubated for 8 (Vero) or 13 (C6/36) days to select for preferentially replicating viruses. Deep sequencing of recovered virus populations showed an expected purging of stop codons (Supplementary Figure 1a) and identified 3 highly enriched substitutions in Vero cells (Q350L, T397S and a synonymous substitution ACA to AGG at position 416) and 2 highly enriched substitutions (K316Q and S461G) in C6/36 cells (Fig 1d). A synonymous substitution selected in mammalian cells likely to have hitchhiked with the other two non-synonymous beneficial substitutions. The strong selection of a small number of substitutions, along with a reduction in codon variants (Supplementary Figure 1b), indicates a bottleneck, which appears to be stronger in Vero cells (mutation frequency of Q350L/T397S/R416R approximately 0.95, and fewer low frequency peaks) than in C6/36 cells (mutation frequency of K316Q/S461G approximately 0.8, and more lower frequency peaks) (Fig 1d and Supplementary data 1 and 2). However, as our goal was not to systematically quantify the effects of all substitutions, but rather to select for specific host-adaptive substitutions, we proceeded with the analysis of enriched substitutions.

The deep sequencing data were also analyzed for differential selection in mosquito compared to mammalian cells. The Supplementary Data 1 file contains a complete step-by-step data analysis of the DMS data using *dms_tools2* (see also https://github.com/jbloomlab/ZIKV_DMS_w_Khromykh). For residues 316 and 461, a variety of amino acids variants were found in the mosquito cell-selected virus population, but in the mammalian cell-selected virus population only the wild type (WT) residues were found at these positions (K316 and S461, Supplementary data 1). For residues 350 and 397, a variety of amino acids variants were found in the mammalian cell-selected virus population, but only the WT residues were found in the mosquito cell-selected virus population (Q350 and T397, Supplementary data 1). This reflects the flexibility that is allowed at these sites in a specific environment (e.g. Vero cells), which is not allowed at the respective sites in the other environment (e.g. C6/36 cells).

To examine whether viruses containing individual codon variants or viruses containing combinations of codon variants were selected in CPER cDNA-transfected cells, the virus populations selected in C6/36 or Vero cells were further passaged at low multiplicity of infection (MOI=0.01) in C36/36 or Vero cells, respectively. The results showed that passaged viruses contained a combination of corresponding mutations, i.e. C6/36 cell-selected virus contained combination of two variants, 316Q and 461G, while Vero cell-

selected virus contained combination of three variants, 350L, 397S, and R416R (Supplementary Figure 2a).

### Validation of identified substitutions.

To further elucidate the roles of the aforementioned amino acid substitutions individually and in combinations, individual mutant viruses 350L and 397S (mammalian-specific mutations), 316Q and 461G (mosquito-specific mutations), as well as double mutants, 350L/397S and 316Q/461G, were generated. Immuno-plaque assay (iPA) in Vero cells showed that 350L, 397S, and 350L/397S mutant viruses produced plaques similar in size to the WT virus, while 316Q, 461G, and 316Q/461G mutant viruses produced smaller plaques (Fig 2a, Supplementary Figure 2b). Mammalian-specific 350L/397S mutant virus grew to higher titers than WT virus in A549 (Fig 2c) and IFNAR$^{-/-}$ MEF (Supplementary Figure 2c) and replicated with similar efficiency to WT virus in Vero cells (Fig 2b) and in C6/36 cells (Fig 2e). Further analyses of RNA replication of mammalian-specific mutants in Vero cells by qRT-PCR revealed that only the double mutation 350L/397S provided significant advantage in RNA replication (Supplementary Figure 2e), indicating that selection in mammalian cells favoured virus with more efficiently replicating RNA.

The mosquito-specific 316Q/461G virus replicated significantly less efficiently in Vero cells (Fig 2b), human epithelial cell line A549 (Fig 2c), and human extravillous trophoblast placental cells HTR8 (Fig 2d), and did not produce detectable replication in the interferon α/β receptor-deficient mouse embryonic fibroblasts (IFNAR$^{-/-}$ MEFs) (Supplementary Figure 2c). The mosquito-specific 316Q/461G virus also induced lower cytopathicity in Vero cells compared to WT virus, while mammalian-specific 350L and 350L/397S viruses induced higher cytopathicity than WT virus (Supplementary Figure 2d). Despite the consistent growth defect observed for the mosquito-specific 316Q/461G virus in all mammalian cell lines tested, this virus replicated with efficiencies similar to WT virus in C6/36 and Aag2 mosquito cell lines (Fig 2e and Supplementary Figure 2f). Notably, while combined 316Q/461G substitutions were stable, a single K316Q substitution was not, giving rise to an additional S66L mutation when the virus was grown in either mammalian or mosquito cell lines (see Supplementary Notes). Virus infectivity assays showed that the defect in virus replication for mosquito-specific viruses in mammalian (Vero) cells was not due to a defect in their ability to infect these cells (Supplementary Figure 3a). Virus binding assay also failed to show any difference in binding to Vero (and C6/36) cells between WT and 316Q/461G viruses (Supplementary Figure 3b). The above observations indicate that these substitutions affect virus assembly/formation and/or secretion in mammalian cells.

### The 316Q/461G virus exhibits temperature-dependent defect in replication in mammalian cells and decreased thermal stability.

Mosquito cells grow at a lower temperature (28°C) than mammalian cells (37°C). The 316Q/461G virus secreted similar levels of infectious virus particles compared with WT virus when grown in Vero cells at 28°C, whereas at 37°C the 316Q/461G virus produced significantly less secreted infectious virus particles than WT virus (Fig 2f). ZIKV particles were reported to have higher thermal stability than other flaviviruses such as DENV[21]. Virus particles produced in C6/36 cells were treated at temperatures ranging from 30°C to 45°C

for 5 hours, and virus infectivity was determined in Vero cells by iPA. The 316Q/461G virus displayed significantly lower thermal stability compared to the WT virus or the 316Q and 461G viruses (Fig 2g).

### The 316Q/461G virus is defective in secretion/spread of infectious virus particles in mammalian cells.

We next tested whether the reduced levels of secreted infectious virus particles for the 316Q/461G virus at 37°C was caused by a defect in either the intracellular accumulation or secretion of virion proteins. Immunoprecipitation analysis with anti-E antibodies of $^{35}$S-labelled infected Vero or C6/36 cells showed that secretion (but not intracellular accumulation of virion proteins) was severely inhibited for the 316Q/461G virus in Vero cells but not in C6/36 cells (Fig 2h). Quantification of virion proteins by ELISA with anti-E antibodies and of viral RNA by qRT-PCR at 3 days post infection (dpi) showed significantly less virion proteins and viral RNA in the culture fluid of Vero cells infected with 316Q/461G virus compared with WT virus (Fig 2i,k). Although intracellular viral E protein levels were somewhat lower for 316Q/461G than for WT virus infection in Vero cells (Fig 2j), intracellular viral RNA levels were not significantly different between the two viruses (Fig 2l). Immunofluorescence analysis of infected Vero and C6/36 cells with anti-E antibody showed that the E protein of 316Q/461G virus was retained in perinuclear regions in Vero cells, but not in C6/36 cells (Fig 3a). In contrast, WT viral E protein was evenly distributed and extended beyond the perinuclear region in both cell lines (Fig 3a). Immuno-staining of infected cells incubated in liquid media (as opposed to solid media used for standard titrations by iPA) after infection further confirmed a defect in the release/spread of 316Q/461G virus in Vero, but not in C6/36 cells (Supplementary Figure 3c,d).

### Transmission electron microscopy of infected cells demonstrate a defect in virion formation for the 316Q/461G virus in mammalian cells.

To further identify steps in the assembly/formation and/or secretion of virions potentially affected by the 316Q/461G substitutions, transmission electron microscopy of infected cells was performed. Compared to C6/36 infections, visibly less virions were seen in Vero cells infected with 316Q/461G virus than with WT virus (Fig 3b), despite the presence of clear markers of infection in the form of induced convoluting membrane structures (CM) (Supplementary Figure 4a). Quantification of virion particles in Vero cells showed significantly less virions per section (n=30) in cells infected with the 316Q/461G virus (Supplementary Figure 4b), with 14 out of 30 sections (46%) showing no virions at all despite clear signs of infection in all imaged sections. The results thus illustrate that 316Q/461G substitutions primarily affect virion formation or stability at 37°C in mammalian cells.

### Molecular analysis and modeling of S461G and K316Q substitutions.

In the heterodimeric mature ZIKV structure, the S461 residue in E is in contact with M protein (Fig 4a). The protein contact analysis by Molecular Operating Environment (MOE) software with default contact threshold settings (Supplementary Table 3) identified contacts of S461 with six residues in M (highlighted in pink in Fig. 4b). The S461G substitution eliminates 3 of these contacts (Fig 4c; Supplementary Table 4), one of which is E24 residue of M previously reported to make a polar contact with S461[22]. Free energy calculations of

the overall E-M interaction in the heterodimer revealed that of all the possible substitutions, G provides the largest change in free energy (Supplementary Table 5) and thus causes the lowest interaction strength between E and M.

Modelling of ZIKV pr protein (using the low pH DENV heterodimeric structure) revealed a potential salt bridge between D57 of pr and K316 of E (Fig 4d, enlarged bottom right). We also noted a complementary interface between the negatively charged underside of pr and the positively charged region (containing K316) surrounding the fusion loop on the E dimer (Supplementary Figure 5a). Patch analysis showed that the K316 residue is key for the formation of a positive patch on E (Supplementary Figure 5b). The K316Q substitution would remove any potential salt bridge and/or the positive patch (Supplementary Figure 5b), which would lead to reduced interaction between pr and E. The S66L mutation that developed in the Q316 virus may compensate by providing increased pr-E interaction via a potential hydrophobic interaction with A47 of pr (Fig 4d, bottom left with L replacing the wildtype S residue).

## The 316Q/461G virus replicates inefficiently and causes less disruption to the development of human brain organoids.

We next assessed the ability of the WT and 316Q/461G viruses to infect and replicate in induced pluripotent stem cell (iPSC)-derived human brain organoids. The 316Q/461G virus replicated less efficiently than WT virus (Fig 5a) and did not induce severe growth retardation of the brain organoids when compared with WT virus infection (Fig 5b,c). In organoids 4 and 18 dpi, detection of virus with antibodies against the ZIKV E protein revealed abundant WT virus in SOX2 (Fig 5d) and BRN2 (Fig 5e) positive neuronal progenitor cells, consistent with previous reports[23,24]. The prevalence of 316Q/461G virus in these progenitor cells was found to be much lower (Fig 5d,e), as expected from the lower viral titers (Fig 5a). The 316Q/461G virus-infected organoids, in most respects, developed similarly to the mock-infected organoids; both showed a clearly delineated outer layer of MAP2-expressing neurons (Fig 5f), whereas WT virus-infected organoids display cortical disorganization at day 18 post infection (Fig 5f). Previous reports indicated that ZIKV virus infection leads to premature differentiation, reduced proliferation and increased cell death in organoids[25,26]. In agreement with these data we observed an early increase in TBR1-expressing cells (Fig 5g), a decreased number of cells expressing the proliferation marker Ki67 (Fig 5h) and an increased number of cells expressing the apoptosis marker activated caspase 3 (Fig 5i) in WT virus-infected organoids. These changes were not observed in organoids infected with the 316Q/461G virus, which exhibited TBR1, Ki67 and activated caspase 3 staining comparable to mock infected organoids (Fig 5g,h,i). Viral RNA replication and changes in expression of marker genes were independently confirmed by qRT-PCR from organoids collected at 18 dpi (Supplementary Figure 6).

## The 316Q/461G virus is highly attenuated in IFNAR$^{-/-}$ mice but induces complete protection against a lethal challenge.

IFNAR1$^{-/-}$ C57BL/6 mice were infected with $10^4$ FFU/mouse of the 316Q/461G or WT viruses by subcutaneous (s.c.) injection, and viremia was assessed daily for 7 dpi (Fig 6a). Infection with WT virus produced a viremia, as observed in our previous studies[12,27], while

no viremia was detected in 316Q/461G-infected mice (Fig 6b). Mouse sera was collected on 21 dpi (Fig 6a), with $PRNT_{50}$ titer between 1:2560–1:5120 detected in WT virus-infected mice, and between 1:640–1:1280 in mice receiving the 316Q/461G virus (Fig 6c). On 37 dpi, mice were challenged s.c. with a lethal dose ($10^3$ $CCID_{50}$) of the African ZIKV-MR766 strain[12] (Fig 6a). ZIKV-MR766 viraemia was detected in the naïve group from day 2 to day 7 post-challenge, while WT and 316Q/461G virus immunized groups showed no viraemia (Fig 6d). Minimal (<5%) to no weight loss was observed in mice from the latter groups (Fig 6e), and all mice survived challenge (Fig 6f). In the naïve group, significant weight loss was recorded up to 7 days post-challenge (Fig 6e), and all mice succumbed to ZIKV-MR766 infection by day 7 post-challenge (Fig 6f). The 316Q/461G virus was thus highly effective in eliciting protective immune responses in $IFNAR^{-/-}$ mice.

The 316Q/461G virions showed similar structural morphology to WT virus by electron microscopy (Supplementary Figure 7a) and both viruses showed similar capacities to be neutralized by anti-ZIKV monoclonal antibodies (Supplementary Figure 7b). In addition, while ZIKV-specific monoclonal antibodies (Supplementary Figure 7c) and human serum (Supplementary Figure 7d) were able to enhance infection of WT virus in human K562 cells, such antibody-dependent enhancement (ADE) was significantly less apparent for the 316Q/461G virus (Supplementary Figure 7c,d). The attenuated replication of 316Q/461G virus in mammalian cells was thus maintained even in ADE settings. Furthermore, both substitutions were stably maintained after 5 passages in either C6/36 or Vero cells (Supplementary Figure 7e). These observations and the ability to grow the virus to high yields in mosquito cells all support the vaccine potential of the 316Q/461G virus. In addition, lack of replication in $IFNAR^{-/-}$ cells and undetectable viraemia in $IFNAR^{-/-}$ mice illustrates effective virus attenuation even when the innate antiviral response is missing; an important safety feature when considering immunization of immunocompromised and pregnant individuals.

## Discussion

Herein we have used DMS methodology and selection in mosquito and mammalian cells to identify two substitutions in the ZIKV E protein, K316Q and S461G, that together strongly favoured virus replication in mosquito cells. A mutant virus 316Q/461G containing these substitutions was subsequently found to be highly attenuated for growth in mammalian cells, including human brain organoids and $IFNAR^{-/-}$ mice.

After exposure to the acidic pH of the trans-Golgi (TGN), the prME arrangement on the flavivirus surface transitions from a heterotrimer to a “flatten” heterodimer[28–30]. This conformational change was proposed to be caused by the tightening of the unfolded portion of pr (termed ‘drawstring’), pulling down the folded head of pr along with the interacting domain II of E, and potentiating/spring-loading the heterodimer structure[31]. The S461 residue on E forms an interaction with M in the heterodimeric[21] (Fig 4a,b) (but not the heterotrimeric) structure and is implicated in the formation and stability of prM-E heterodimers[32,33]. The reduced number of virus particles (Supplementary Figure 4b) and the reduced M-E interaction (Fig 4c), suggests the S461G substitution destabilizes the prM-E

heterodimers. This may promote virion disassembly during the energy intensive structural rearrangements that occur in the transition from heterotrimer to heterodimer.

As a result of the structural rearrangements in the TGN, the furin cleavage site on prM is exposed, allowing cleavage of prM to pr and M. The pr protein remains associated with E (via non-covalent interactions) to prevent exposure of the fusion loop and premature fusion in low pH environment of the TGN[34]. Additionally, pr was also proposed to bind to both E monomers on the heterodimeric structure, and latching them in this spring-loaded conformation[31]. K316 is positioned below the pr domain in the low pH structure[35–37], and our modelling suggests that it may form either a salt bridge (Fig 4d, bottom right) and/or charge interactions with pr (Supplementary Figure 5). The K316Q substitution may thus diminish the non-covalent association between pr and E, thereby potentially promoting release of the latch and premature exposure of the fusion loop. Such activity may initiate premature fusion events during virus transit through TGN. The S66L mutation may partially compensate by providing increased pr-E interactions.

In addition, secreted 316Q/461G particles were also much less thermally stable than WT virus particles (Fig 2g), consistent with diminished E-M and/or E-pr interactions caused by the 316Q/461G substitutions. Co-incidentally, a T267Q mutation in E-DII has recently been reported to render ZIKV thermally unstable, with the disruption of hydrophobic interactions between E and M proposed to be the mechanism[38]. Our results and this observation illustrate the importance of E-M interaction in ensuring high thermal stability of ZIKV virions.

The mosquito cell-selected 316Q and 461G substitutions identified herein have not been found in any arbovirus from genus flaviviruses sequenced to date. However, these amino acids can be found in insect-specific flaviviruses (Supplementary Notes and Supplementary Table 2). Why these substitutions were selected in C6/36 cells remains unclear. Studies of DENV2 suggest there is a reduced capacity to adopt configurations leading to fusion at lower temperatures[31]. Conceivably, the 316Q and 461G substitutions may have been selected because they reduce the energy hurdles for transition from the smooth structure through to the ultimate infectious fusogenic forms at lower temperatures.

In summary, we describe herein the use of a DMS screen to identify residues in the E protein of ZIKV that confer advantage to virus replication in primate or mosquito cells. The 316Q/461G virus, which incorporates two substitutions selected in mosquito cells, replicated efficiently in mosquito cells at 28°C, but grew poorly in mammalian cells at 37°C. The 316Q/461G virus also showed potential as an attenuated vaccine, showing no detectable viraemia in IFNAR$^{-/-}$ mice, but provided complete protection against lethal ZIKV challenge (Fig 6). Our findings illustrate the power of DMS for identifying amino acids that are key to regulating virus replication in specific host environments and for discovery of virus strains with vaccine potential.

## Methods

### Generation and deep sequencing of DMS DNA amplicon for C-terminal 204 codons of E gene.

Codon mutagenesis was performed following the method previously described[39], stopping at the step of joining PCR. The forward and reverse mutagenesis primers are described in Supplementary Table 1. The forward mutagenesis primers pool is used in combination with the Natal2_R primer (5' CTTTCACGGGGTGTCCAATTAGCTCTGAAG 3'), while the reverse mutagenesis primers pool is used in combination with the Natal1_F primer (5' AGTTGTTGATCTGTGTGAATCAGAC 3'). The left fragment pool (indicated by the blue lines in Fig 1a) ranges from 1880–2489 bp, and the right fragment pool (indicated by red lines in Fig 1a) ranges from 1367–1979 bp. Successful amplification of these fragments were confirmed on DNA gel electrophoresis which appeared as short smears. The left and right fragment pools were gel purified, and then subjected to joining PCR (or anneal and extension PCR). Briefly, a 325.5 ng amount of purified left fragment pool (representing approximately $1.38 \times 10^{11}$ copies of dsDNA) and 242 ng amount of right fragment pool (representing approximately $1.33 \times 10^{11}$ copies of dsDNA) were added into the joining PCR reaction using the PrimeSTAR GXL DNA polymerase (Takara Bio, USA). Joining PCR was performed for 30 cycles, using a touchdown program (annealing temperature set from 66°C to 36°C over 30 cycles) and the fused fragment purified by gel purification. The DNA amplicon fragment containing the mutagenized region was then PCR amplified, using the Natal1_F (5' AGTTGTTGATCTGTGTGAATCAGAC) and Natal2_R (5' CTTTCACGGGGTGTCCAATTAGCTCTGAAG) primers, and including 147 ng of joining PCR fragment (approximately $3.53 \times 10^{10}$ copies of dsDNA) as template. The extent of mutagenesis observed in the DNA amplicon fragment was analyzed as described in Supplementary Data 1 (see also https://github.com/jbloomlab/ZIKV_DMS_w_Khromykh). Barcoded subamplicon sequencing was previously described (in reference[4], and in the dms_tools2 documentation - https://jbloomlab.github.io/dms_tools2/bcsubamp.html), except that the DMS region in ZIKV E-DIII stem-anchor was divided into two barcoded subamplicons labelled 'E1' and 'E2' (Supplementary Figure 8a). Briefly, the method is divided into two stages (PCR1 and PCR2, Supplementary Figure 8b). In PCR1, E1 and E2 fragments are amplified using the primers E1_F (5' CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNcgcctgaaaatggataaacttagattg), E1_R (5' GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNtgccactcctgtgccagtg), and E2_F (5' CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNggggagaagaagatcacccac), E2_R (5' GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNcaccgagcaccccacatc). The partial adaptors are indicated in upper-case, random barcodes are indicated by the 8 Ns, and the gene-specific sequences are indicated by the lower case letters. PCR1 products are then gel purified and diluted to reduce barcode diversity in PCR2 (Supplementary Figure 8c). PCR2 reaction was performed using the primers UniversalRnd2for (5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC), and index01Rnd2rev (5' CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCAGACGTGTGCTCTT

CCGATCT), which completes the adaptor sequences. Sequencing was performed on Illumina MiSeq v3 2×300 bp.

### Computational analysis of deep sequencing data

The deep sequencing data were analysed using version 2.2.dev6 of the dms_tools2 software, which is available at http://jbloomlab.github.io/dms_tools2. Supplementary Data 1 contains a HTML rendered file, which contains a Jupyter notebook with plots detailing the deep sequencing data analysis as generated using the dms_tools2 software. All code and input data are provided as Supplementary Data 2 that performs all steps of the analysis beginning with downloading the FASTQ files from the Sequence Read Archive. The actual code in these files is also available on GitHub at https://github.com/jbloomlab/ZIKV_DMS_w_Khromykh. As the paired-end reads for sequencing of the mutated region were separated into two subamplicons (Supplementary Figure 8), the combined (e.g. double or triple) mutations across the entire mutated region could not be determined.

### Generation of infectious ZIKV DMS DNA library by CPER and selection in Vero and C6/36 cells.

The above DMS DNA amplicon was then used in equimolar amount with other ZIKV cDNA fragments and a linker fragment containing either mammalian CMV[12,16,17,19,20] or insect OpIE2[18] promoters in a CPER reaction to produce infectious DNA library as previously described[12]. The CPER-assembled infectious DNA library containing the CMV promoter was then transfected into Vero cells while the CPER-assembled infectious DNA library containing the OpIE2 promoter was transfected into C6/36 cells. Selection of beneficial mutations was carried out by incubating transfected Vero cells for 8 days and transfected C6/36 cells for 13 days. Secreted viruses were collected at these times, viral RNA purified by Trizol LS (Sigma Aldrich, USA), E1 and E2 barcoded subamplicons generated by RT-PCR (as in Supplementary Fig 8), deep sequenced, and deep sequencing data analysed as above. Differential selection analysis of mutations in C6/36 versus Vero cells was then performed using dms2_batch_diffsel program within the dms_tools2 package.

### Cell culture

K562, HTR-8 and C6/36 cells were maintained in RPMI 1640 media. A549 cells were maintained in F-12K media. Vero and IFNAR$^{-/-}$ MEF cells were maintained in DMEM media. All cells were supplemented with 10% FCS, 50 units/mL of penicillin and 50 μg/mL of streptomycin. Mammalian cells were cultured in a 37°C incubator with 5% $CO_2$, while C6/36 mosquito cells were cultured at 28°C.

### Immuno-plaque assay (iPA)

Approximately $2\times10^4$ cells/well for Vero cells and $1\times10^5$ cells/well for C6/36 cells were seeded in a 96-well plate and incubate for 12 h up to 90%−100% confluent. Virus samples were diluted in 10-fold serial dilutions in DMEM (with antibiotic and 5% of FCS) and 25 μL from each dilution was added into each well containing the cells, and incubated at 37°C. Two-hour post infection, 175 μL of overlay media (2X M199 medium, 5% FCS, 50 units/mL penicillin, 50 μg/mL streptomycin, 2% carboxymethyl cellulose) was added. Forty-eight

hours after infection, the overlay was removed and 100 μL of ice-cold 80% acetone in PBS was added and incubated at −20°C for 1 hour to fix the cells, and then air dried. For the antibody probing, the plate was first blocked for 30 min at 37°C with 50 μL of blocking solution (Pierce™ Clear Milk Blocking Buffer, Thermo Scientific, USA), followed by the addition of primary 50 μL of mAb 4G2 (mouse anti-E) for 1h at 37°C. Plates were then washed six times with PBS/T (PBS containing 0.05% Tween 20). Secondary antibody (IRDye 800CW Goat anti-Mouse, LI-COR, USA) was added (25 μL/well) and incubated for 1h at 37°C. Plates were washed six times, air dried and then kept in the dark at room temperature until ready for imaging. Plates were scanned using the LI-COR Biosciences Odyssey Infrared Imaging System using the following specifications (channel=800, intensity=auto, focal length=3 mm, resolution=42 μm). Viral titers from iPA are expressed as focus forming units (FFU) per mL (FFU/mL).

### Viral growth kinetics

All mammalian cells were seeded at $5\times10^5$ cells per well in 6-well plates, while C6/36 cells were seeded at $1\times10^6$ cells per well in 6-well plates. Cells were then infected with viruses at the multiplicity of infection (MOI) indicated in the figures. Briefly, 200 μl of virus inoculum diluted to the required MOI was added onto each well of the 6-well plates and incubated at 37 °C with 5% $CO_2$ (mammalian cells) or at 28°C (mosquito cells) for 1 hour with rocking at every 15 minutes interval. After infection, the cell monolayers were washed three times with additive-free DMEM/RPMI. Thereafter, 3 mL of media (depending on cell types; see Cells section) was added to each well. At the time points indicated in the figures, a 150 μL culture supernatant sample was harvested and stored at −80°C. Samples were then titered by iPA as described above.

### Antigen-capture ELISA

Nunc MaxiSorp flat-bottom 96-well plates (Nunc, Roskilde, Denmark) were coated with fusion peptide specific mAb BJ-6E6 (125 ng per well) at 4 °C overnight in PBS. Culture supernatant or cell lysates harvested in NP40-lysis buffer (CAT# FNN0021, Thermo Fisher Scientific) were diluted in two-fold serial dilutions in blocking buffer (0.05 M Tris–HCl pH 8.0, 1 mM EDTA, 0.15 M NaCl, 0.05%, v/v Tween 20, 0.2%, w/v casein) and then incubated in the coated plates for 2 hours at 37°C. This is followed by addition of a human anti-E-DIII mAb ZV-67 used at a concentration of 0.31 μg/mL in blocking buffer, and incubated at room temperature for 1 hour, and detected with a HRP-conjugated rabbit anti-human IgG antibody (Dako), used at a 1:3000 dilution in blocking buffer. Plates are washed two times with PBS/T (PBS supplemented with 0.05% Tween 20) between all steps. ABTS substrate solution (0.02% (w/v) ABTS, 0.06% (v/v) $H_2O_2$) was added and wells were read on a UV plate reader at 405 nm after 1 h (Labsystems Multiscan EX Type 355, Pathtec, VIC, Australia). ELISA threshold was set at mean plus two times the standard deviation of mock.

### Real-time RT-PCR

RNA purification was performed using TRI Reagent (Sigma Aldrich, USA), or from culture supernatant using NucleoSpin RNA Virus (Macherey-Nagel). Purified RNA samples were added with primers specific for ZIKV RNA (Fwd 5’ CTTGGAACTCCTCTACAGCCAC, Rev 5’ GAGTAGAACTCCAGGGCCGA), TBP-1 (Fwd 5’

CAGCTTCGGAGAGTTCTGGG, Rev 5' CAAACCGCTTGGGATTATATTCGG), RPL11 (Fwd 5' CGGTCTGGACTTCTACGTTGTGCT, Rev 5' CCACTTCATGGCGTCTTCCTTGG), EFTA (Fwd 5' TGTTGATGCTGGCTTTGTTC, Rev 5' TGGATGGCTCCAGATATTCC), BRN2 (Fwd 5' CGGCGGATCAAACTGGGATTT, Rev 5' TTGCGCTGCGATCTTGTCTAT), TBR1 (Fwd 5' ACAATTTTCCTGACTCCAAGG, Rev 5' ACTGTGACGAAGCTCAGAGAC), or Ki67 (Fwd 5' CTGACCCTGATGAGAGTGAGGGA, Rev 5' GGGCTTCTCCCCTTTTGAGAG), and using the Luna Universal One-Step RT-qPCR Kit (NEB) according to the manufacturers' instructions. Real-time RT-PCR was performed using the QuantStudio 6 Flex Real-Time PCR instrument (Thermo Fisher Scientific).

### Transmission Electron Microscopy of infected cells

Cell pellets were fixed in 3% glutaraldehyde in PBS for 2 hours at room temperature, washed with 0.1 M cacodylate buffer (CDB) and incubated in 1% $OsO_4$ (in 0.1 M CDB) for 40 min at room temperature. Samples were then washed once with 0.1 M CDB and once in 80% acetone for a further incubation overnight at 4°C in 2% uranyl acetate UAc/80% acetone. The next day serial dehydrating and resin infiltration steps were performed as follows: 2 × 10 min with 80% acetone, 2 × 10 min with 90% acetone, 3 × 20 min with 100% acetone, 1 × 90 min with 50% Epon/50% acetone, 1 × 90 min with 75% Epon/25% acetone, 1 × 90 min with 100%Epon. Epon was replaced with fresh 100% Epon with polymerization accelerator BDMA and embedded at 65°C for 72 hours. Resin blocks were then trimmed using a DiATOME Ultra Diamond Knife on a Leica EM UC7 - Ultramicrotome and 50nm thickness sections were obtained and mounted on EM-copper grids with formvar/carbon coating. Sections were post-stained in 4% UAc in water and Reynolds' lead citrate for 5 min each and then processed for TEM-imaging using a FEI Tecnai F20 S/TEM electron microscope or Talos L120C TEM.

### Heat treatment of virus particles

To analyse thermal sensitivity of viruses, approximately $10^6$ FFU of each virus was subjected to a thermal gradient treatment from 30°C to 45°C for 5 hours using a thermocycler (Bio-Rad T100 Thermal Cycler, USA), after which samples were immediately titered by iPA on Vero cells. Relative infectivity was determined by calculating the ratio of number of FFU in heat-treated versus the respective non-heat treated viruses (kept at 4°C for 5h). Three independent experiments were conducted.

### Radio-immunoprecipitation assay

Vero or C6/36 cells were seeded at a cell density of $5\times10^5$ or $1\times10^6$ cells per well of a 6-well plate, respectively. The next day, cells were infected at an MOI 10 for Vero cells, and MOI 1 for C6/36 cells with ZIKV-Natal, ZIKV-316Q/461G or uninfected. At 6 dpi, C6/36 cells were incubated in DMEM supplemented with 50 μg/mL Actinomycin D for 6 hours, followed by starvation media (DMEM, L-methionine-free, L-cysteine-free, Life Technologies) supplemented with 50 μg/mL Actinomycin D for 30 mins to deplete endogenous methionine and cysteine, and then supplemented with $^{35}$S-L-methionine and $^{35}$S-L-cysteine (EasyTag™ EXPRESS35S Protein Labeling Mix, Perkin Elmer) to a final concentration of 200 μCi/mL. Cells were incubated for a further 24 hours, and then

supernatant was harvested and clarified by centrifugation, while cell lysates were harvested in 500 μL NP-40 lysis buffer (50 mM Tris, pH 7.4, 250 mM NaCl, 5 mM EDTA, 1% Nonidet™ P40) supplemented with cOmplete™ Protease Inhibitor Cocktail (Sigma Aldrich) and lysed by passing through a 29G insulin needle (Becton Dickinson). Vero samples follow the same procedure except that ZIKV-Natal-infected cells were treated at 2 dpi, while ZIKV-316Q/461G and uninfected cells were treated at 3 dpi. All samples are stored at −80°C.

Immunoprecipitation was performed using Dynabeads™ Protein A (Life Technologies). Briefly, for each sample 50 μL of Dynabeads Protein A were loaded with 10 μg of purified 6B6C-1 mAb (anti-E) in a 1.5 mL tube by incubation on a rotary mixer at 4°C. After 2 hours, unbound antibodies were removed by washing 1X with PBS/T, followed by incubation with 1 mL of supernatant samples or 250 μL of cell lysate samples, and allowed to incubate at 4°C on the rotary mixer for a further 2 hours. Dynabeads were washed 2X with PBS/T and then proteins eluted by incubating in 30 μL of NuPAGE LDS Sample Buffer (Life Technologies) at 70°C for 5 mins. Twenty-five μL of each sample were then loaded into each lane of a NuPAGE 4–12% Bis-Tris Protein Gels, 1.0 mm, 15-well (Life Technologies), and resolved by SDS-PAGE. Proteins were then transferred onto a nitrocellulose membrane using the iBlot 2 Dry Blotting System (Life Technologies) according to the manufacturer's instructions, and then exposed onto a storage phosphor screen (GE Healthcare Life Sciences) for 3 days and then imaged on the Typhoon FLA 7000 (GE Healthcare Life Sciences).

### Culture of iPSC-derived human brain organoids

Generation of human induced pluripotent stem cells (iPSC) was previously described[40] with the following modifications:

Cell line Genea022[41] (Genea Biocells, San Diego, CA) were adapted to single cell format and cultured in mTESR (CAT# 85850, Stem Cell Technologies) on matrigel (CAT# 354234, Corning) coated plates. On day 0 of organoid culture the iPSCs were dissociated with StemPro Accutase (CAT# A1110501, GIBCO), seeded at 15,000 cells per well in mTESR plus ROCK inhibitor (CAT# 130-103-922, Miltenyi Biotech) in a round–bottom, ultra–low attachment 96-well plate (Corning) and then spun at 330×g for 5 minutes at room temperature to form aggregates. On day 1, iPSCs were differentiated into neural progenitor cells (NPCs) and were maintained in 200 μL Knockout-serum replacement (KSR) with dual SMAD (KSR+SMAD) inhibition medium containing 80% w/v DMEM/F12 (CAT# 11320–033, GIBCO), 20% w/v knockout serum replacement (CAT# 10828028, GIBCO), 1% w/v Glutamax (CAT# 35050–061, GIBCO), 1% w/v MEM–NEAA (CAT# 11140–050, GIBCO), 0.2% w/v 10,000 U/mL Penicillin–Streptomycin (CAT# 15140–122, GIBCO), 2 μM dorsomorphin (CAT# P5499, Sigma-Aldrich) and A83 (CAT# RDS293910, In Vitro Technologies) for 5 days with daily media replacements. From days 6 to 14, KSR+SMAD inhibition media was replaced by neural induction media with containing 95% w/v DMEM/F12, 1% w/v Glutamax, 1% w/v MEM–NEAA, 1% w/v 10,000 U/mL Penicillin–Streptomycin, 1% w/v N2 (CAT# 17502–048, Life Technologies), 1% w/v 1 μg/ml heparin (CAT# H3149, Sigma-Aldrich), 1 μM CHIR99021 (CAT# 130-106-539, Miltenyi Biotech)

and SB431542 (CAT# 130-106-543, Miltenyi Biotech) to induce differentiation into cerebral organoids. From day 15 onwards, differentiated cerebral organoids were cultured in neural differentiation (ND) media comprising 1:1 DMEM/F12 (CAT# 11320–033, GIBCO) and Neurobasal media (CAT# 21103–049, GIBCO), 1% w/v Glutamax, 1% w/v MEM–NEAA, 1% w/v 10,000 U/mL Penicillin–Streptomycin, 1% w/v N2, 2% w/v B27 (CAT# 17054–044, Life Technologies), 0.02% w/v Insulin (CAT# I9278, Sigma-Aldrich) and 0.09% w/v 2-Mercaptoethanol (CAT# 21985–023, GIBCO).

### Infection of iPSC-derived human brain organoids

Cerebral organoids at day 15 were utilized for viral infection. A virus inoculum titer of $10^4$ FFU per 50 µl of ZIKV-Natal or ZIKV-316Q/461G was added onto a single organoid containing well of a round–bottom, ultra–low attachment 96–well plate and incubated at 37°C for 24 hour. Each ZIKV–infected organoid was then transferred into a single well of a 24-well plate (Corning) containing 500 µl of ND media.

To determine viral growth kinetics in infected organoids, at the indicated timepoints, 160 µl of culture supernatant was harvested from each well, and then replaced with the same amount of fresh culture media. Harvested culture supernatants were titered by iPA. Three infected organoids per virus were used. Infected organoids were imaged by dark-field microscopy with 4× magnification using Nikon Eclipse TE200 Inverted Microscope, and organoid dimensions were measured using ImageJ.

For immunofluorescence analysis, infected organoids were harvested at 4 and 18 dpi, fixed overnight in 4% paraformaldehyde, and left in 30% sucrose solution for 48h before being embedded in OCT (Leica). The CryoStar NX70 Cryostat (Thermo Fisher Scientific) was used to prepare 10 µm cryosections, which were then stained for organoid markers and co-stained with the 4G2 mAb (anti ZIKV E). Antibodies against organoid markers are anti-rabbit SOX2 (CST, CAT# D9B8N), BRN2 (CST, CAT# D2C1L), Caspase-3 (CST, CAT# D175), Ki67 (CST, CAT# D3B5), MAP-2 (Thermo Fisher Scientific, CAT# PA5–17646) and TBR1 (Abcam, CAT# ab31940). All slides were were imaged at 22°C on a Leica SP8 confocal laser scanning microscope (ANFF-Q).

### Molecular modelling and simulation

The Molecular Operating Environment (MOE) 2016.08 (Chemical Computing Group, Montreal Inc., Canada) was used for molecular computational chemistry. Energy minimization of the K316Q or S461G mutated structure, interaction analysis of the residues, free energy calculations of the single amino acid substitutions, and protein surface patch analysis were conducted in the setting of the Amber10:EHT force field and the generalized Born/volume integral (GB/VI) implicit solvent model. The default MOE settings were used for other detailed parameters.

### Antibody-Dependent Enhancement (ADE) assay

Twenty-five µl of 2-fold serially diluted mAbs or serum were incubated with an equal volume of virus at 37°C for 1 h, at MOI of 3. K562 cells adjusted to a concentration of $5 \times 10^4$ cells/50 µl in serum-free RPMI 1640 medium were added to the mAb-virus mixture in

1.5mL micro centrifuge tube and incubated at 37°C for 2 h with gentle tap every 15 minutes. Cells were then washed twice with 1 mL of serum-free RPMI 1640 to remove unbound virus and mAb. Cells were incubated in RPMI 1640 containing 2% FCS. Culture supernatants were harvested at 48 hours post infection, clarified and viral titers quantified by iPA on Vero cells. Three independent experiments were performed.

### Plaque reduction neutralization test

Vero cells were seeded in 96-well plates as according to the iPA section above. Sera from infected IFNAR1$^{-/-}$ C57BL/6 mice were inactivated at 56°C for 30 minutes. Two-fold serial dilutions of the serum samples (from 1:5 to 1:10240 in DMEM containing 4% FCS, 50 units/mL penicillin and 50 μg/mL streptomycin) or mAb 4G2, ZV67 and C8 were incubated with 50–100 FFU of ZIKV-MR766, ZIKV-Natal (WT) or ZIKV-316Q/461G for 20 min at room temperature and then at 37°C for 1 hour. Subsequently, each mixture (100uL) was added to Vero cell monolayer and incubated at 37°C. After 2 hours, the inoculum was removed and overlay media added. The plates were incubated at 37°C with 5% of $CO_2$ atmosphere for 24h. The cells were fixed and probed as described above (iPA section), with the exception that PRNT plates of mouse serum samples were probed using human mAb ZV-67 (anti-E) as primary antibody and IRDye 800CW Goat anti-Human as secondary antibody. Relative infectivity was determined by tabulating the ratio of number of FFU in each serum/mAb dilution with the no-antibody control for each virus.

### Viral cytotoxicity assay

Vero cells were seeded into 96 well plates at the density of $2 \times 10^4$ cells per well. Serial 0.5-$Log_{10}$ dilutions of the of the viruses with virus concentration from $2 \times 10^6$ to 2 FFU/ml were prepared and 100 ul of each dilution were used to infect the cells at 24h after seeding; 8 wells of cells were left uninfected (mock) to be used as a control. Infection was performed by incubation with the inoculums for 24 h, then inoculums were replaced with 100 ul/well DMEM supplemented with 2% FCS. At 5 dpi cytopathic effect was determined based on ATP levels in the cells using Viral ToxGlo Assay (Promega, USA) according to the manufacturer's instructions. Luminescence was measured on DTX880 Multimode Detector (Beckman Coulter, USA). Percentage of survival was determined as a percentage of luciferase activity (luminescence value) in infected cells comparing to uninfected control and CPE was calculated as 100% - (% of survival). Experiment was performed in duplicates. Data was fitted into sigmodal curves using GraphPad Prism v7.0.

### Immunofluorescence analysis of infected cells.

Vero or C6/36 cells were seeded on glass coverslips in a 24-well plate at a density of $1\times10^4$ and $1\times10^5$ cells per well, respectively. Cells were then infected with ZIKV-Natal or ZIKV-316Q/461G at MOI 1 and incubated at 37°C (Vero) or 28°C (C6/36). After one hour, cells were washed and replaced with fresh media, and then incubated at 37°C (Vero) or 28°C (C6/36). Both ZIKV-Natal and ZIKV-316Q/461G-infected C6/36 cells were fixed and stained at 5 dpi. ZIKV-Natal –infected Vero cells were fixed and stained at 3 dpi, and ZIKV-316Q/461G-infected Vero cells were fixed at 4 dpi. Fixing and staining procedures are as follows: glass coverslips were removed and submerged in 100% acetone at −20°C for 5 minutes, and then allowed to air dry completely. Coverslips were then blocked with 250 μL

blocking solution (Pierce™ Clear Milk Blocking Buffer, Thermo Scientific, USA) for 1 hour at room temperature, before incubation with 4G2 (anti-E) diluted in PBS/T for a further 1 hour. Coverslips were then washed 3 times with PBS/T and incubated with goat anti-mouse IgG Alexa Fluor 488 secondary antibody (Life Technologies) diluted in PBS/T for a final 1 hour. Coverslips were then washed 3 times with PBS/T and then mounted onto microscope glass slides with ProLong™ Diamond Antifade Mountant with DAPI (Life Technologies). Images were taken using the Zeiss LSM 710 confocal microscope.

### Cryo-Electron Microscopy of secreted virus particles

ZIKV-316Q/461G virions were prepared for TEM analysis as described previously[42]. Briefly, virus was precipitated from clarified culture supernatants using PEG 8000, resuspended in NTE buffer (10 mM Tris pH 8, 120 mM NaCl, 1 mM EDTA) before ultracentrifugation through a 20% sucrose cushion and a final purification step on a 10–40% potassium tartrate gradient. The virus band was harvested and buffer exchanged into NTE and final preparations (4 μl) were applied to a freshly glow-discharged 400-mesh holey carbon-coated copper grids (Proscitech). Grids were blotted for 7 s, in 100% relative humidity before plunge-freezing in liquid ethane using a Vitrobot (FEI). Cryo-EM images were acquired at 300 kV with an FEI Tecnai F30 G2 microscope, equipped with a K2 summit direct electron detector (Gatan).

### Viral infectivity assay

To assess the infectivity of viral mutants in different cell types, all virus stocks were first titered on iPA using C6/36 monolayer and normalized to approximately $10^5$ FFU/mL and separately aliquoted. Virus aliquots were then simultaneously titered on iPA using C6/36 or Vero cell monolayers in the same experiment. Three independent experiments were conducted.

### Viral spread assay

Vero or C6/36 cells were seeded at a cell density of $5\times10^5$ or $1\times10^6$ cells per well of a 6-well plate, respectively. The next day cells were infected at MOI 0.0001 with ZIKV-Natal or ZIKV-316Q/461G for 1 hour, and then washed and incubated in their respective culture media. Plates were fixed and stained, and then imaged at the at different timepoints using the same procedures as described in the iPA section above.

### Mouse experiments

Ten-weeks old IFNAR1$^{-/-}$ C57BL/6 male mice were infected via subcutaneous (s.c.) injection with ZIKV-Natal (WT), ZIKV-316Q/461G or uninfected (naïve group), and viraemia was monitored daily for 7 dpi. At 21 dpi, blood samples were harvested by tail bleed to determine the neutralizing antibody response levels against virulent ZIKV-MR766. At 37 days post-vaccination, mice were challenged with a lethal dose of $10^3$ $CCID_{50}$/animal of ZIKV-MR766 and viraemia was determined daily for 7 days post-challenge. Weight loss and mortality rate were measured for up to 21 days post-challenge.

### Ethics statement

All mouse experiments were conducted in accordance with the “Australian code for the care and use of animals for scientific purposes” as defined by the National Health and Medical Research Council of Australia. Animal experiments and associated statistical treatments were reviewed and approved by the QIMR Berghofer Medical Research Institute animal ethics committee (P2195).

### Data Availability

Deep sequencing data are deposited on the Sequence Read Archive under PRJNA449413. Data files are also provided as Supplementary Data 1 and 2. Software suite for DMS data analysis can be found on https://github.com/jbloomlab/dms_tools2. The actual code used to run the suite in this study is at https://github.com/jbloomlab/ZIKV_DMS_w_Khromykh

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia grant APP1144950. AAK and ASuh are Research Fellows with the NHMRC. EN was supported in part by the Daiichi Sankyo Foundation of Life Science, Japan. We thank Rob Sullivan from the Queensland Brain Institute Microscopy and Histology Facility for the help with preparation and imaging of the brain organoid slides. Organoid confocal microscopy work was performed in part at the Queensland node of the Australian National Fabrication Facility (ANFF-Q)—a company established under the National Collaborative Research Infrastructure Strategy to provide nano and microfabrication facilities for Australia’s researchers.

## References

1. Musso D & Gubler DJ Zika Virus. Clin Microbiol Rev 29, 487–524, doi:10.1128/CMR.00072-15 (2016). [PubMed: 27029595]
2. Ashenberg O, Padmakumar J, Doud MB & Bloom JD Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. PLoS Pathog 13, e1006288, doi: 10.1371/journal.ppat.1006288 (2017). [PubMed: 28346537]
3. Dingens AS, Haddox HK, Overbaugh J & Bloom JD Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. Cell Host Microbe 21, 777–787 e774, doi:10.1016/j.chom. 2017.05.003 (2017). [PubMed: 28579254]
4. Doud MB & Bloom JD Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses 8, doi:10.3390/v8060155 (2016).
5. Doud MB, Hensley SE & Bloom JD Complete mapping of viral escape from neutralizing antibodies. PLoS Pathog 13, e1006271, doi:10.1371/journal.ppat.1006271 (2017). [PubMed: 28288189]
6. Pierson TC & Diamond MS The emergence of Zika virus and its new clinical syndromes. Nature 560, 573–581, doi:10.1038/s41586-018-0446-y (2018). [PubMed: 30158602]
7. Mittal R et al. Zika Virus: An Emerging Global Health Threat. Front Cell Infect Microbiol 7, 486, doi:10.3389/fcimb.2017.00486 (2017). [PubMed: 29276699]
8. Carod-Artal FJ Neurological complications of Zika virus infection. Expert Rev Anti Infect Ther 16, 399–410, doi:10.1080/14787210.2018.1466702 (2018). [PubMed: 29668332]
9. Qi H et al. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. PLoS Pathog 10, e1004064, doi:10.1371/journal.ppat. 1004064 (2014). [PubMed: 24722365]

10. Wu NC et al. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Sci Rep 4, 4942, doi:10.1038/srep04942 (2014). [PubMed: 24820965]
11. Mlakar J et al. Zika Virus Associated with Microcephaly. N Engl J Med 374, 951–958, doi: 10.1056/NEJMoa1600651 (2016). [PubMed: 26862926]
12. Setoh YX et al. De Novo Generation and Characterization of New Zika Virus Isolate Using Sequence Data from a Microcephaly Case. mSphere 2, doi:10.1128/mSphereDirect.00190-17 (2017).
13. Fontes-Garfias CR et al. Functional Analysis of Glycosylation of Zika Virus Envelope Protein. Cell Rep 21, 1180–1190, doi:10.1016/j.celrep.2017.10.016 (2017). [PubMed: 29091758]
14. Gong D et al. High-Throughput Fitness Profiling of Zika Virus E Protein Reveals Different Roles for Glycosylation during Infection of Mammalian and Mosquito Cells. iScience 1, 97–111, doi: 10.1016/j.isci.2018.02.005 (2018). [PubMed: 30227960]
15. Bloom JD Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics 16, 168, doi:10.1186/s12859-015-0590-4 (2015). [PubMed: 25990960]
16. Edmonds J et al. A novel bacterium-free method for generation of flavivirus infectious DNA by circular polymerase extension reaction allows accurate recapitulation of viral heterogeneity. J Virol 87, 2367–2372, doi:10.1128/JVI.03162-12 (2013). [PubMed: 23236063]
17. Amarilla AA et al. Chimeric viruses between Rocio and West Nile: the role for Rocio prM-E proteins in virulence and inhibition of interferon-alpha/beta signaling. Sci Rep 7, 44642, doi: 10.1038/srep44642 (2017). [PubMed: 28317911]
18. Piyasena TBH et al. Infectious DNAs derived from insect-specific flavivirus genomes enable identification of pre- and post-entry host restrictions in vertebrate cells. Sci Rep 7, 2940, doi: 10.1038/s41598-017-03120-1 (2017). [PubMed: 28592864]
19. Setoh YX et al. Helicase Domain of West Nile Virus NS3 Protein Plays a Role in Inhibition of Type I Interferon Signalling. Viruses 9, doi:10.3390/v9110326 (2017).
20. Setoh YX et al. Systematic analysis of viral genes responsible for differential virulence between American and Australian West Nile virus strains. J Gen Virol 96, 1297–1308, doi:10.1099/vir. 0.000069 (2015). [PubMed: 25626681]
21. Kostyuchenko VA et al. Structure of the thermally stable Zika virus. Nature 533, 425–428, doi: 10.1038/nature17994 (2016). [PubMed: 27093288]
22. Sevvana M et al. Refinement and Analysis of the Mature Zika Virus Cryo-EM Structure at 3.1 A Resolution. Structure 26, 1169–1177 e1163, doi:10.1016/j.str.2018.05.006 (2018). [PubMed: 29958768]
23. Watanabe M et al. Self-Organized Cerebral Organoids with Human-Specific Features Predict Effective Drugs to Combat Zika Virus Infection. Cell Rep 21, 517–532, doi:10.1016/j.celrep. 2017.09.047 (2017). [PubMed: 29020636]
24. Qian X et al. Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. Cell 165, 1238–1254, doi:10.1016/j.cell.2016.04.032 (2016). [PubMed: 27118425]
25. Sutarjono B Can we better understand how Zika leads to microcephaly? A systematic review of the effects of the Zika virus on human brain organoids. J Infect Dis, doi:10.1093/infdis/jiy572 (2018).
26. Gabriel E et al. Recent Zika Virus Isolates Induce Premature Differentiation of Neural Progenitors in Human Brain Organoids. Cell Stem Cell 20, 397–406 e395, doi:10.1016/j.stem.2016.12.005 (2017). [PubMed: 28132835]
27. Prow NA et al. A vaccinia-based single vector construct multi-pathogen vaccine protects against both Zika and chikungunya viruses. Nat Commun 9, 1230, doi:10.1038/s41467-018-03662-6 (2018). [PubMed: 29581442]
28. Rey FA, Stiasny K, Vaney MC, Dellarole M & Heinz FX The bright and the dark side of human antibody responses to flaviviruses: lessons for vaccine design. EMBO Rep 19, 206–224, doi: 10.15252/embr.201745302 (2018). [PubMed: 29282215]
29. Mukhopadhyay S, Kuhn RJ & Rossmann MG A structural perspective of the flavivirus life cycle. Nat Rev Microbiol 3, 13–22, doi:10.1038/nrmicro1067 (2005). [PubMed: 15608696]
30. Sirohi D & Kuhn RJ Zika Virus Structure, Maturation, and Receptors. J Infect Dis 216, S935–S944, doi:10.1093/infdis/jix515 (2017). [PubMed: 29267925]

31. Zhang X et al. Cryo-EM structure of the mature dengue virus at 3.5-A resolution. Nat Struct Mol Biol 20, 105–110, doi:10.1038/nsmb.2463 (2013). [PubMed: 23241927]
32. Allison SL, Stadler K, Mandl CW, Kunz C & Heinz FX Synthesis and secretion of recombinant tick-borne encephalitis virus protein E in soluble and particulate form. J Virol 69, 5816–5820 (1995). [PubMed: 7637027]
33. Allison SL, Stiasny K, Stadler K, Mandl CW & Heinz FX Mapping of functional elements in the stem-anchor region of tick-borne encephalitis virus envelope protein E. J Virol 73, 5605–5612 (1999). [PubMed: 10364309]
34. Yu IM et al. Association of the pr peptides with dengue virus at acidic pH blocks membrane fusion. J Virol 83, 12101–12107, doi:10.1128/JVI.01637-09 (2009). [PubMed: 19759134]
35. Prasad VM et al. Structure of the immature Zika virus at 9 A resolution. Nat Struct Mol Biol 24, 184–186, doi:10.1038/nsmb.3352 (2017). [PubMed: 28067914]
36. Yu IM et al. Structure of the immature dengue virus at low pH primes proteolytic maturation. Science 319, 1834–1837, doi:10.1126/science.1153264 (2008). [PubMed: 18369148]
37. Li L et al. The flavivirus precursor membrane-envelope protein complex: structure and maturation. Science 319, 1830–1834, doi:10.1126/science.1153263 (2008). [PubMed: 18369147]
38. Xie DY et al. A single residue in the alphaB helix of the E protein is critical for Zika virus thermostability. Emerg Microbes Infect 7, 5, doi:10.1038/s41426-017-0006-9 (2018). [PubMed: 29362473]
39. Bloom JD An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol 31, 1956–1978, doi:10.1093/molbev/msu173 (2014). [PubMed: 24859245]
40. Lancaster MA et al. Cerebral organoids model human brain development and microcephaly. Nature 501, 373–379, doi:10.1038/nature12517 (2013). [PubMed: 23995685]
41. Dumevska B, Bosman A, McKernan R, Schmidt U & Peura T Derivation of human embryonic stem cell line Genea022. Stem Cell Res 16, 472–475, doi:10.1016/j.scr.2016.02.011 (2016). [PubMed: 27346017]
42. McLean BJ et al. A novel insect-specific flavivirus replicates only in Aedes-derived cells and persists at high prevalence in wild Aedes vigilax populations in Sydney, Australia. Virology 486, 272–283, doi:10.1016/j.virol.2015.07.021 (2015). [PubMed: 26519596]

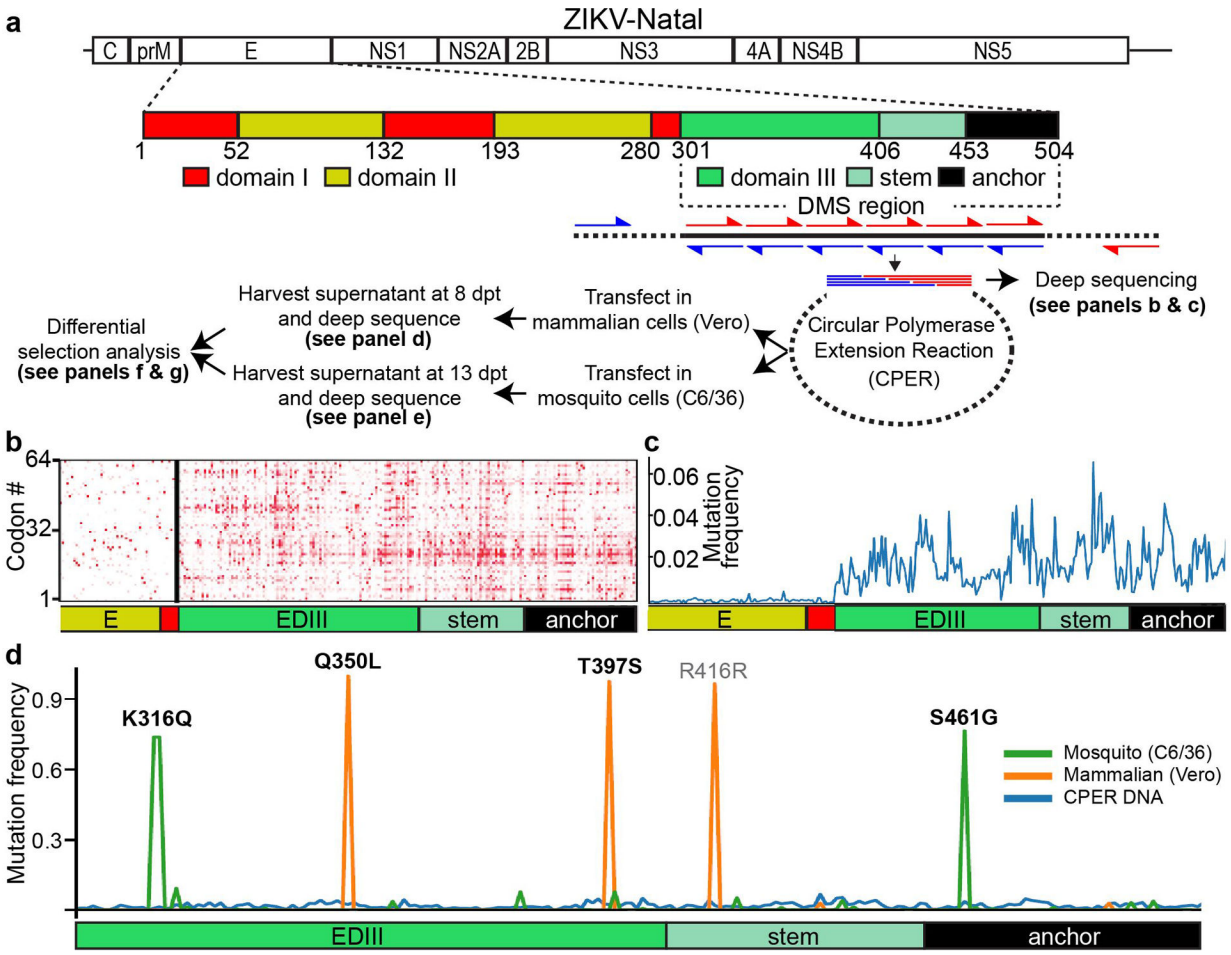


**Figure 1.**

Deep mutational scanning (DMS) of C-terminal region of ZIKV E protein. **a)** Schematic of ZIKV E protein and workflow of DMS screening for the C-terminal 204 codons of ZIKV E protein. Amplicons generated using forward (red arrows) and reverse (blue arrows) mutagenesis primers are combined with other ZIKV cDNA fragments to generate infectious cDNA library by Circular Polymerase Extension Reaction (CPER). The CPER cDNA library is then transfected into mammalian (Vero) or mosquito (C6/36) cells, recovered viruses harvested in culture supernatants, deep sequenced and analysed to identify preferentially selected viral mutants. **b)** Graphical representation of deep sequencing data showing codon variant heatmap, and **c)** Mutation frequency of cDNA amplicon library illustrating the extent of mutagenesis in the E-DIII stem-anchor region. Mutation frequency is defined as the percentage at which codon mutants are detected in sequenced molecules at each site. For example, at each site, if mutations are present in 100% of sequenced molecules, the frequency is 1. If it is not present in any sequenced molecules, the frequency is 0. **d)** Mutation frequency of CPER cDNA library, mammalian (Vero) cells-selected virus population and mosquito (C6/36) cells-selected virus population. Identities of preferentially selected amino acid mutations are indicated above each peak.

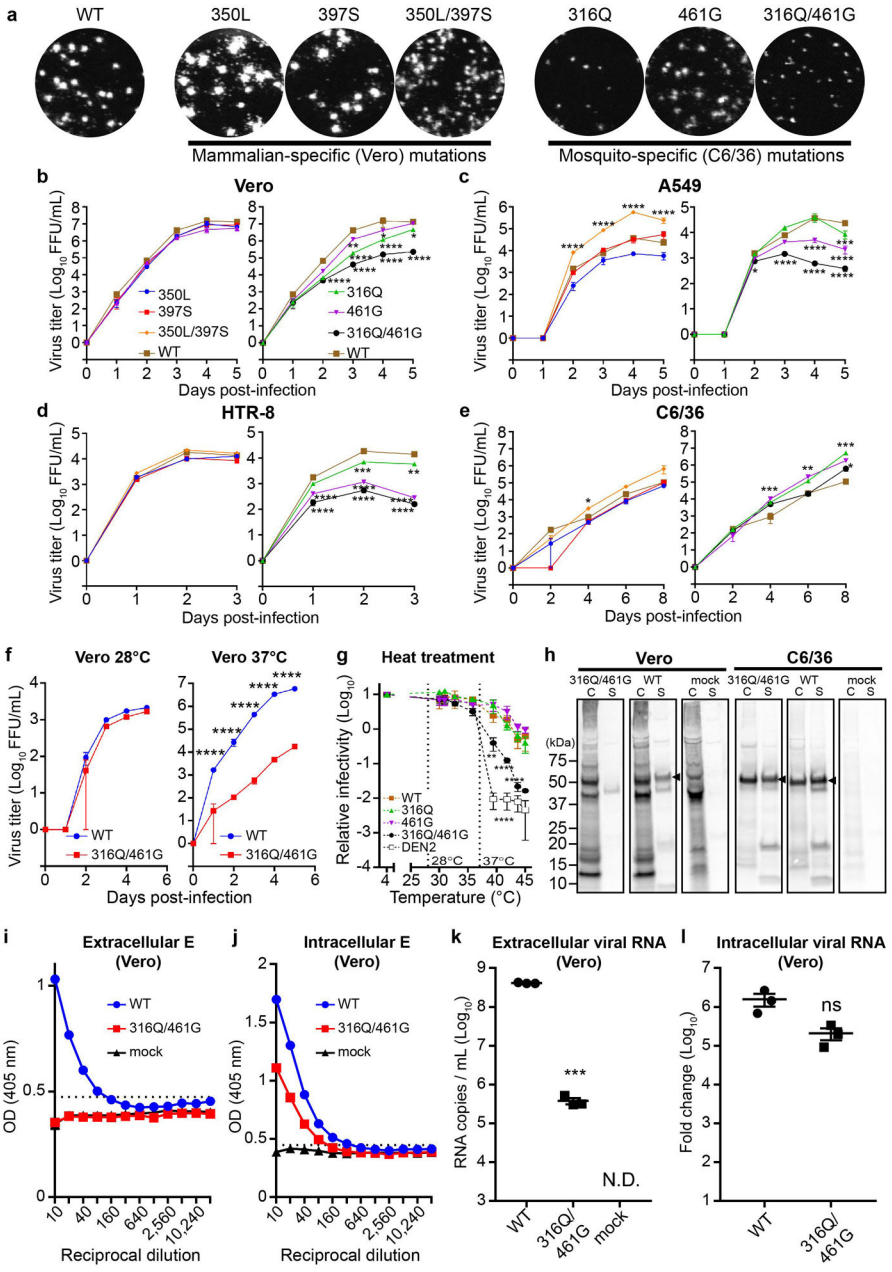


**Figure 2.**

In vitro characterization of ZIKV mutants. **a)** Immuno-plaque assay (iPA) of WT and mutant viruses in Vero cells. **b-e**) Growth kinetics of WT and mutant viruses infected at a multiplicity of infection (MOI) of 0.1 in **b)** Vero, **c)** A549, **d)** HTR-8, and **e)** C6/36 cells. **f)** Growth kinetics of Vero cells infected with WT or 316Q/461G mutant at MOI 0.1, and incubated at 28°C or 37°C. **g)** WT and mutant viruses, as well as control dengue serotype 2 (strain TSV-01) virus were subjected to heat treatment at the indicated temperatures for 5 hours, and then titered. All culture supernatants above were harvested at their indicated time points after infection and virus titres determined by iPA in Vero cells. n=3 independent experiments for all assays, and statistical analysis was performed by two-way ANOVA with

Tukey's multiple comparisons test against WT virus. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. Mean ± standard error of the mean. Limit of detection for iPA is 1.6 $Log_{10}$FFU/mL. **h)** Immunoprecipitation with anti-E antibody (6B6C-1) of $^{35}$S-labelled cell lysates (C) or supernatants (S) from Vero or C6/36 cells infected with WT or 316Q/461G mutant viruses. Envelope protein indicated by the arrowhead. Antigen-capture ELISA detection of **i)** extracellular and **j)** intracellular envelope protein. Dotted lines show the limit of detection, n=3 independent experiments. Detection of **k)** extracellular and **l)** intracellular viral RNA by qRT-PCR. N.D. = not detected. Mean ± standard error of the mean, n=3 independent experiments. Statistical analysis for qRT-PCR data was performed using unpaired t-test, two-tailed. *** $P = 0.0005$.

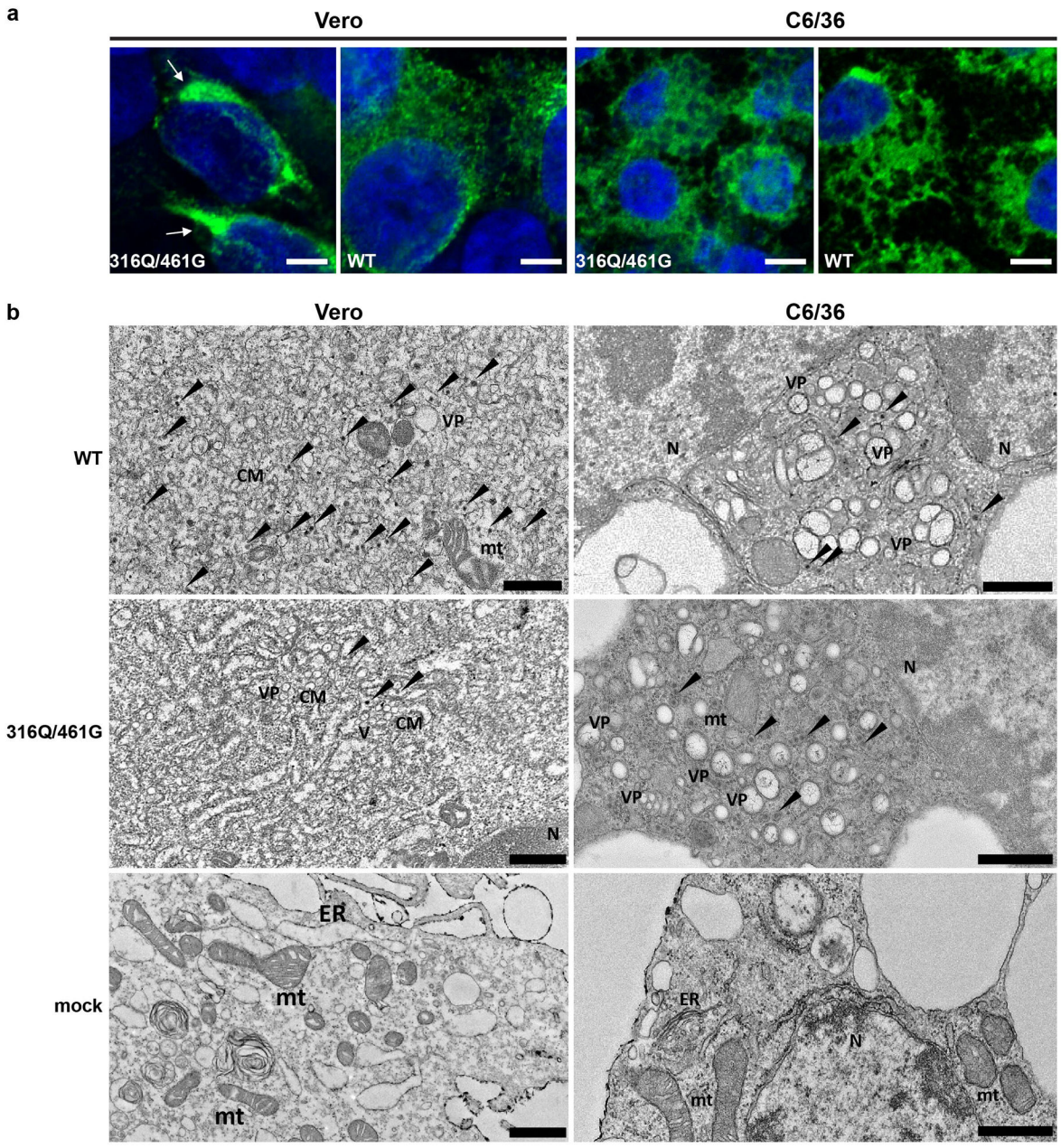


**Figure 3.**

Immunofluorescence and transmission electron microscopy of infected C6/36 or Vero cells. a) Immunofluorescence microscopy of Vero or C6/36 cells infected with WT or 316Q/461G mutant, and immuno-stained with 4G2 anti-E antibody (green channel), and DAPI (blue channel). Scale bars are 5 μm. Representative images of n=3 biologically independent samples. b) Transmission electron microscopy of C6/36 or Vero cells infected with WT, 316Q/461G mutant, or uninfected (mock). Virus particles are indicated by arrowheads. CM – convoluted membranes, VP – vesicle packets, V –vesicles, N - nucleus, mt – mitochondria, ER – endoplasmic reticulum. Scale bars in C6/36 slides are 200 nm, and scale bars in Vero slides are 500 nm. n=3 biologically independent samples.

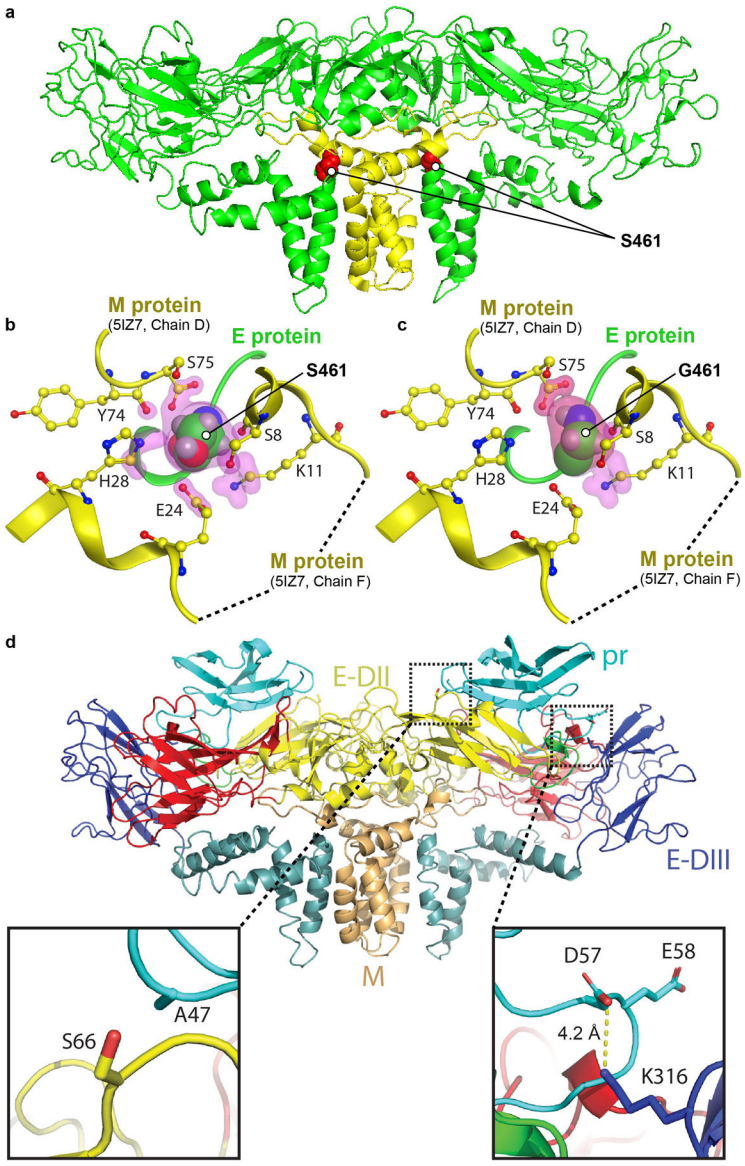


**Figure 4.**

Molecular analysis and modelling of the S461G and K316 mutations. **a)** Cryo-EM structure of H/PF/3013 strain ZIKV E dimer (PDB accession number: 5IZ7) showing the position of the S461G mutation (red spheres) in the mature structure. E and M proteins represented by green and yellow ribbon structures, respectively. **b)** Residue S461 sits in a pocket -in contact with six M protein residues; Y74, S75 (from chain D in M protein) and S8, K11, E24, H28 (from chain F in M protein). Contact atoms are surrounded by pink highlights. Protein contact analysis by Molecular Operating Environment (MOE) software package (using 5IZ7). **c)** The S461G substitution results in only 3 residues in contact with G461. The S461G mutation was introduced on the 5IZ7 structure followed by energy minimization and protein contact analysis by MOE. **d)** Homology model of ZIKV Natal pr protein (based on DENV structure PDB:3C5X and generated using Swiss-Model) fitted to the cryo-EM structure for low pH immature DENV (EMD-5006), showing a potential salt bridge forming

between D57 in pr and K316 in E-DIII (right inset). A S66L mutation can potentially result in hydrophobic interaction between L66 in E-DII and A47 in pr (left inset).

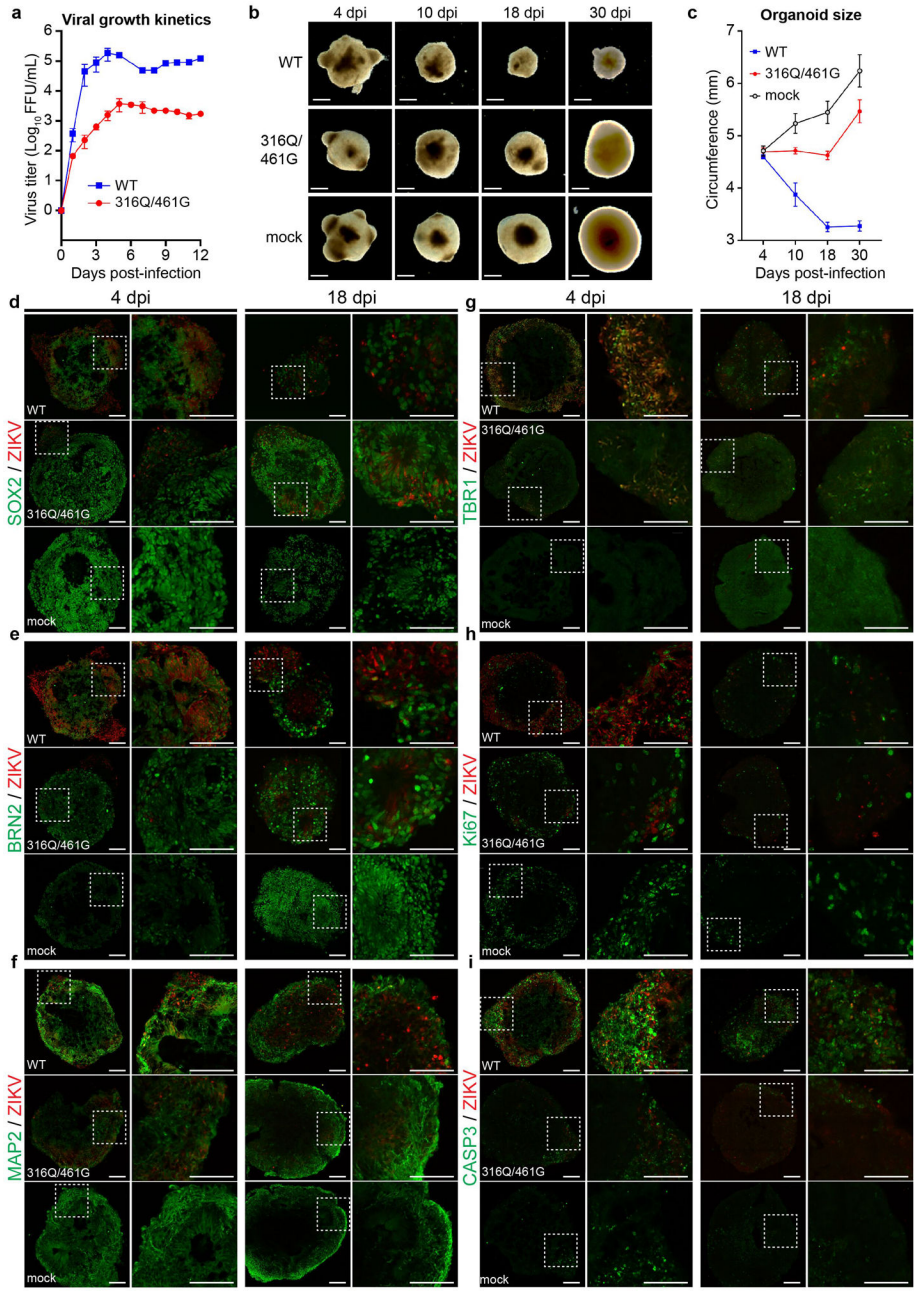


**Figure 5.**

Infection of iPSC-derived human brain organoids. **a)** Growth kinetics of WT or 316Q/461G viruses in iPSC-derived human brain organoids. Three different organoids per virus were infected. Mean ± standard deviation. **b)** Morphology of iPSC-derived human brain organoids infected with WT virus, 316Q/461G mutant virus, or uninfected visualized by light microscopy. Scale bars are 0.355 mm. Representative images of n=3 biologically independent samples per timepoint. **c)** Quantification of organoid sizes after infection. Circumference of four organoids per group were measured. Mean ± standard deviation. Immunohistochemistry (IHC) analysis of iPSC-derived human brain organoids co-stained for ZIKV E protein and **d)** SOX2, **e)** BRN2, **f)** MAP2, **g)** TBR1, **h)** Ki67, **i)** CASP3 at the 4

and 18 dpi timepoints. Scale bars are 100 μm. Paired images represent the original image on the left, with the zoom-in of the inset (dotted box) on the right. Representative images of n=3 biologically independent samples per timepoint.

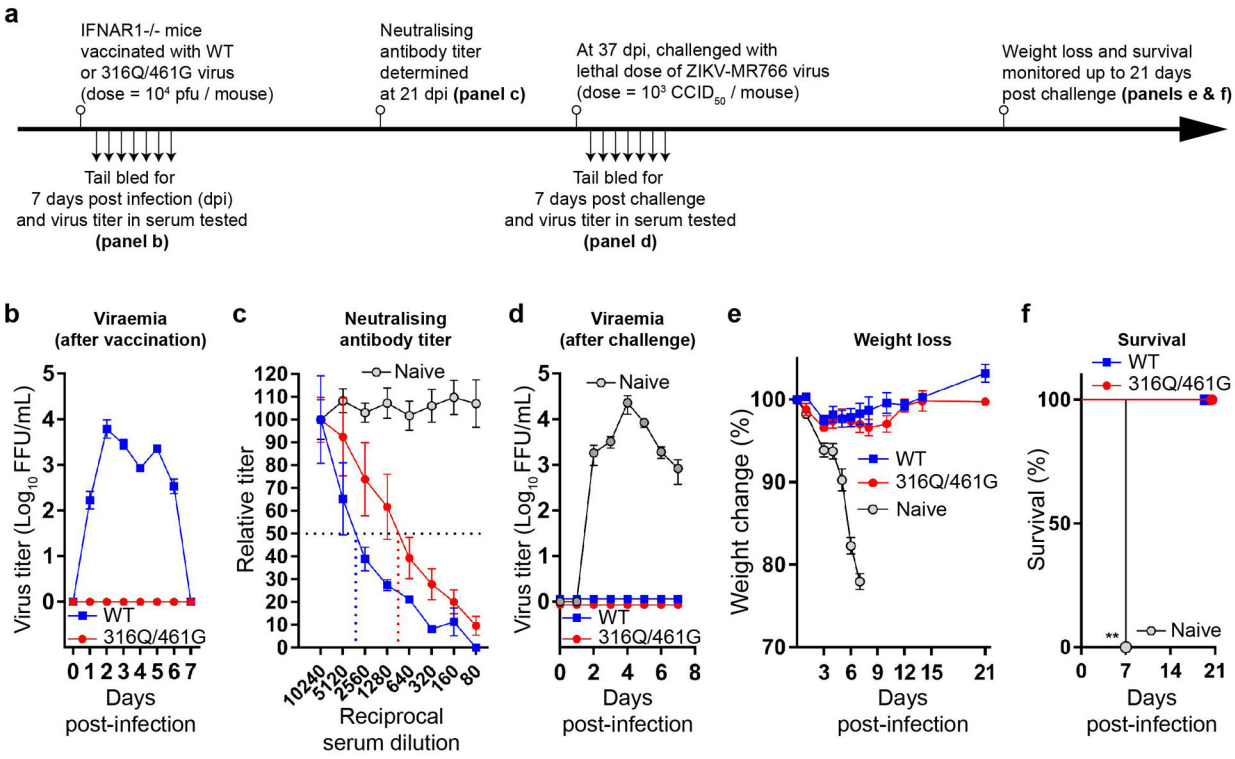


**Figure 6.**

316G/461Q mutant as an attenuated ZIKV vaccine candidate. **a)** Timeline of mouse experiment. **b)** Viraemia in IFNAR1$^{-/-}$ C57BL/6 mice vaccinated with $10^4$ FFU/mouse of WT (n=5) or 316Q/461G virus (n=5). Daily tail bleeds were conducted up to 7 days after vaccination and virus titres in the collected sera were determined by iPA in Vero cells. Mean ± standard error of the mean. **c)** Neutralising antibody titers against ZIKV-MR766 virus in the mouse sera collected at 21 days post-vaccinations with WT virus (n=5), 316Q/461G mutant virus (n=5) or naïve (n=4). The neutralizing antibody titres were determined by plaque reduction neutralization test (PRNT) on Vero cells. Mean ± standard error of the mean. **d)** Viraemia in IFNAR1$^{-/-}$ C57BL/6 mice vaccinated with WT virus (n=4) or 316Q/461G virus (n=5) or naïve (n=4) and challenged with $10^3$ $CCID_{50}$/mouse of ZIKV-MR766. Daily tail bleeds were conducted up to 7 days post-challenge, and virus titres in collected mouse sera were determined by iPA in Vero cells. Mean ± standard error of the mean. **e)** Percentage body weight loss of IFNAR1$^{-/-}$ C57BL/6 mice vaccinated with WT (n=4) or 316Q/461G (n=5) or naïve (n=4) was determined up to 21 days after challenge with ZIKV-MR766. Mean ± standard error of the mean. **f)** Survival of IFNAR1$^{-/-}$ C57BL/6 mice vaccinated with WT (n=4) or 316Q/461G (n=5) or naïve (n=4) and challenged with ZIKV-MR766. Statistical analysis performed by Log-rank (Mantel-Cox) test, two-sided, compared to WT. ** P = 0.0025.